Transboundary and Emerging Diseases

Transboundary and Emerging Diseases

ORIGINAL ARTICLE

# Foot-and-Mouth Disease Virus Serotype O Phylodynamics: Genetic Variability Associated with Epidemiological Factors in Pakistan

B. P. Brito[1,5], A. M. Perez[1,2], S. M. Jamal[3,4,*], G. J. Belsham[4], S. J. Pauszek[5], Z. Ahmed[5] and L. L. Rodriguez[5]

[1] Center for Animal Diseases Modeling and Surveillance, University of California-Davis, Davis, CA, USA
[2] CONICET/CECSA, Facultad de Ciencias Veterinarias, Universidad Nacional de Rosario, Argentina
[3] National Veterinary Laboratory, Islamabad, Pakistan
[4] National Veterinary Institute, Technical University of Denmark, Lindholm, Denmark
[5] USDA/ARS Foreign Animal Disease Research Unit, Plum Island Animal Disease Center, P.O. Box 848, Greeport NY 11944, USA

**Keywords:**
Foot-and-mouth disease; Pakistan; epidemiology; genetic variation; buffalo; cattle

**Correspondence:**
B. P. Brito. Center for Animal Diseases Modeling and Surveillance, University of California-Davis, 1044 Haring Hall, One Shields Avenue, Davis, CA, 95616, USA.
Tel.: +01 (530) 752 0336;
Fax: +01 (530) 752 1618;
E-mail: bbrito@ucdavis.edu

*Present address: Ministry of Commerce, Islamabad, Government of Pakistan.

Received for publication April 30, 2012

doi:10.1111/j.1865-1682.2012.01366.x

## Summary

One of the most challenging aspects of foot-and-mouth disease (FMD) control is the high genetic variability of the FMD virus (FMDV). In endemic settings such as the Indian subcontinent, this variability has resulted in the emergence of pandemic strains that have spread widely and caused devastating outbreaks in disease-free areas. In countries trying to control and eradicate FMD using vaccination strategies, the constantly evolving and wide diversity of field FMDV strains is an obstacle for identifying vaccine strains that are successful in conferring protection against infection with field viruses. Consequently, quantitative knowledge on the factors that are associated with variability of the FMDV is prerequisite for preventing and controlling FMD in the Indian subcontinent. A hierarchical linear model was used to assess the association between time, space, host species and the genetic variability of serotype O FMDV using viruses collected in Pakistan from 2005 to 2011. Significant ($P < 0.05$) amino acid and nucleotide variations were associated with spatial distance, but not with differences in host species, which is consistent with the frequent multi-species infection of this serotype O FMDV. Results from this study will contribute to the understanding of FMDV variability and to the design of FMD control strategies in Pakistan. Viruses sequenced here also provide the earliest reported isolate from the Pan Asia $II^{ANT-10}$ sublineage, which has caused several outbreaks in the Middle East and spread into Europe (Bulgaria) and Africa (Libya).

## Introduction

Foot-and-mouth disease (FMD) is a highly contagious disease of livestock, producing losses that are insidious in endemic settings and devastating in disease-free countries (James and Rushton, 2002; Thompson et al., 2002; Perry and Rich, 2007). FMD is caused by infection with a single-stranded, positive-sense RNA virus, from the *Picornaviridae* family, genus *Aphthovirus*, referred to as FMD virus (FMDV). The FMDV genome (ca. 8 400 nucleotides) includes a single large open reading frame (ORF), which is ~7 000 nucleotides long. It encodes a polyprotein that is processed into the leader (L) protein, four structural proteins: VP4 (1A), VP2 (1B), VP3 (1C) and VP1 (1D), plus the 2A peptide, as well as non-structural proteins from the P2 (2B and 2C) and P3 (3A, $3B_{1-3}$, 3C and 3D) precursors (Sobrino et al., 2001; Grubman and Baxt, 2004; Belsham, 2005; Carrillo et al., 2005). Seven serotypes, referred to as A, O, C, SAT 1, SAT 2, SAT 3 and Asia 1, have been described based on the antigenic

  

characteristics of the virus (Brooksby, 1958). Serotypes O, A and Asia 1 are prevalent among the ruminant population of Pakistan (Klein et al., 2008; Valarcher et al., 2009; Jamal et al., 2010, 2011a,b,c). FMD control and eradication in endemic countries, such as Pakistan, would not only result in increased agricultural productivity, but also will allow for the international trading of animals and animal products. This could ultimately result in improved quality of life for the rural population (Perry and Rich, 2007). In Pakistan, vaccination against FMD is voluntary and performed by few livestock owners. Moreover, effectiveness of vaccines available in the country is uncertain (Jamal et al., 2010).

Many recent FMD outbreaks in FMD-free countries have been caused by viral strains that have originally emerged from the Indian subcontinent (Knowles and Samuel, 2003; Valarcher et al., 2008). This region offers favourable conditions for FMDV emergence, likely due to a combination of insufficient control measures, large susceptible animal populations, and frequent and unrestricted livestock movements. An endemic setting with such characteristics allow for the existence of a wide variety of genetically diverse viruses. In turn, the high mutation rate of FMDV allows the virus to adapt to different environments (Domingo et al., 2003). Genetic variation is the result of mutation and recombination events throughout the virus genome during its replication. When these mutations result in amino acid substitutions within the structural proteins, the antigenic characteristics of the virus may change and consequently allow the virus to evade the host's immune response (Haydon et al., 2001). FMDV antigenic variation has been attributed to amino acid substitutions in major antigenic sites located on exposed regions of the capsid proteins VP1, VP2 and VP3 (Acharya et al., 1989; Haydon et al., 2001; Yoon et al., 2011b). The most studied epitope is located in the G-H loop of the VP1 protein, although epitopes within VP2 and VP3 also have an important role in the antigenic characteristics of the virus (Sobrino et al., 2001). Antigenic variation is a major obstacle for FMD control, because it results in an absence of cross-protection between serotypes and, in many cases, even between strains of the same serotype. In this study, we have quantified the extent to which epidemiological factors contributed to accumulated nucleotide (nt) and amino acid (aa) variation in the serotype O FMDV (FMDV-O) field samples circulating in Pakistan over a period of 6 years (2005–2011). Results presented here will help to gain critical understanding of the epidemiological dynamics of FMD in Pakistan and, ultimately, to design and implement effective control measures in the country that might be mirrored at the global level.

## Materials and Methods

### Data source

A total of 82 nucleotide sequences encoding VP1 from FMDV-O isolates collected from 2005 to 2011 at 22 different locations in Sindh, Punjab and Khyber Pakhtunkhwa Provinces of Pakistan were included in the analysis. Sixty-eight sequences, processed at two laboratories, were obtained from published literature, 36 from the National Veterinary Institute (DTU-VET, Lindholm) in Denmark and 32 from the Institute for Animal Health in Pirbright in the U.K. Additionally, 14 samples from which FMDV was isolated were processed and sequenced (whole capsid coding region, P1) at the Plum Island Animal Disease Center (PIADC) in the USA. Furthermore, P1 sequences were generated from five FMDVs at DTU-VET, Lindholm in Denmark. One P1 sequence was taken from the published literature (Belsham et al., 2011). Thus, a total of 20 P1 sequences were used in the analysis (Table 1).

Epidemiological information associated with each isolate included species (cattle or buffalo), location (lower administrative division) and date of collection. The location was geocoded (latitude and longitude) using the centre of the polygon of the lower administrative division reported at a web-based library (available at http://www.heavens-above.com/). A standard nomenclature was given to all the sequences for a more convenient visualization of figures: the first three text characters correspond to the lower administrative location of sample collection, the following two numbers are the consecutive samples from a particular year, the next two numbers indicate the year and the last capital letter corresponds to the species from which the sample was collected (C = cattle, B = buffalo), that is, sequence KHI1706B is the 17th sample in 2006 collected in Karachi and from a buffalo. The corresponding virus name as previously published is indicated in Table 1.

### Viral RNA detection and sequencing

At PIADC, viral RNA was extracted using a high-throughput method described elsewhere (Pacheco et al., 2010). RT-PCR products were generated using SuperScript®III One-Step RT-PCR System with Platinum® *Taq* High Fidelity (Invitrogen, Grand Island, NY, USA). The amplification primers were designed to amplify the entire P1 region of any FMDV isolate and were obtained from the Animal Plant and Health Inspection Services at PIADC. The RT-PCR products were visualized, purified, sequenced and analysed following procedures described elsewhere (Pauszek et al., 2011). Additional internal sequencing primers specific to the Pakistan isolates were used to obtain the complete P1 and are available from

 

**Table 1.** Foot-and-mouth disease virus serotype O used in the study. [Correction added on date 2 April 2013, after first online publication: Sequence name 'O/SGD/PAK/10/2010[a]' has been corrected to 'O/IBD/PAK/10/2010[a]' and some accession numbers also have been changed in Table 1.]

| Sequence name | Fig. 2 Nomenclature | Sample location | Species affected | Accession number | Reference |
|---|---|---|---|---|---|
| O/PUN/PAK/L181/2005 | SGD0305C | Sargodha | Cattle | HQ439224 | Jamal et al. (2011a) |
| O/PUN/PAK/L149/2005 | GUJ0405B | Kharian, Gujrat | Buffalo | HQ439222 | Jamal et al. (2011a) |
| O/PUN/PAK/L150/2005 | GUJ0505B | Kharian, Gujrat | Buffalo | HQ439223 | Jamal et al. (2011a) |
| O/PUN/PAK/L133/2005 | LHR0705C | Lahore | Cattle | HQ439221 | Jamal et al. (2011a) |
| O/ISL/PAK/L285/2005 | IBD1205B | Islamabad | Buffalo | HQ439226 | Jamal et al. (2011a) |
| O/ISL/PAK/L286/2005 | IBD105B | Islamabad | Buffalo | HQ439227 | Jamal et al. (2011a) |
| O/ISL/PAK/L287/2005 | IBD1305B | Islamabad | Buffalo | HQ439228 | Jamal et al. (2011a) |
| O/PAK/68/2006 | HBD1606B | Hafizabad | Buffalo | FJ798173 | Waheed et al. (2010) |
| O/PUN/PAK/L1370/2009[a] | RWP0909B | Rawalpindi | Buffalo | HQ439215–JX171677[b] | Jamal et al. (2011a) This study |
| O/PAK/1/2008 | BKI0108C | Burki | Cattle | FJ798190 | Waheed et al. (2011) |
| O/PAK/10/2006 | KHI1106B | Landhi Cattle Colony, Karachi | Buffalo | EF494503 | Klein et al. (2008) |
| O/PAK/11/2006 | KHI1706B | Karachi | Buffalo | EF494504 | Klein et al. (2008) |
| O/PAK/12/2006 | KHI1806B | Nagori Cattle Colony, Karachi | Buffalo | EF494505 | Klein et al. (2008) |
| O/PAK/14/2006 | KHI1906B | Nagori Cattle Colony, Karachi | Buffalo | EF494506 | Klein et al. (2008) |
| O/PAK/2/2006 | KHI1206C | Quadriba, Karachi | Cattle | EF494499 | Klein et al. (2008) |
| O/PAK/2/2008 | PTK0208B | Patoki | Buffalo | FJ798191 | Waheed et al. (2011) |
| O/PAK/3/2008 | PTK0308C | Patoki | Cattle | FJ798192 | Waheed et al. (2011) |
| O/PAK/38/2005[c] | KRK0605C | Karak | Cattle | FJ798167 | Waheed et al. (2011) |
| O/PAK/4/2006 | KHI1306C | Landhi Cattle Colony, Karachi | Cattle | EF494500 | Klein et al. (2008) |
| O/PAK/53/2007 | SGD0107C | Sargodha | Cattle | FJ798179 | Waheed et al. (2011) |
| O/PAK/6/2006 | KHI1406B | Landhi Cattle Colony, Karachi | Buffalo | EF494501 | Klein et al. (2008) |
| O/PAK/6/2008 | PWR0408C | Peshawar, KPK | Cattle | FJ798193 | Waheed et al. (2011) |
| O/PAK/61/2006 | SGD0506C | Sargodha | Cattle | FJ798169 | Waheed et al. (2011) |
| O/PAK/66/2007 (B) | LHR0707B | Lahore | Buffalo | FJ798185 | Waheed et al. (2011) |
| O/PAK/68/2007 | TTS0607B | TT Singh | Buffalo | FJ798186 | Waheed et al. (2011) |
| O/PAK/69/2007 | LYH0807B | Layyah | Buffalo | FJ798187 | Waheed et al. (2011) |
| O/PAK/70/2007 | LYH0907B | Layyah | Buffalo | FJ798188 | Waheed et al. (2011) |
| O/PAK/71/2007 | LMT1107C | Lakki Marwat | Cattle | FJ798189 | Waheed et al. (2011) |
| O/PAK/8/2006 | KHI1506B | Landhi Cattle Colony, Karachi | Buffalo | EF494502 | Klein et al. (2008) |
| Pak 23.1_8 | KHI0106C | Karachi | Cattle | EF494490 | Klein et al. (2008) |
| Pak 23.2_8 | KHI0206C | Karachi | Cattle | EF494491 | Klein et al. (2008) |
| Pak 24.1_8 | KHI0306C | Karachi | Cattle | EF494493 | Klein et al. (2008) |
| Pak 5.2_9 | KHI0406B | Karachi | Buffalo | EF494498 | Klein et al. (2008) |
| O/PUN/PAK/L282/2005[c] | LYH1405C | Layyah | Cattle | HQ439225 | Jamal et al. (2011a) |
| O/ISL/PAK/L288/2005[c] | IBD1505C | Islamabad | Cattle | HQ439229 | Jamal et al. (2011a) |
| O/NWF/PAK/L1414/2009 | HAR0409C | CBF, Harichand | Cattle | HQ439212 | Jamal et al. (2011a) |
| O/NWF/PAK/L1415/2009 | HAR0509C | CBF, Harichand | Cattle | HQ439213 | Jamal et al. (2011a) |
| O/NWF/PAK/L1416/2009 | HAR0609C | CBF, Harichand | Cattle | HQ439214 | Jamal et al. (2011a) |
| O/NWF/PAK/L1417/2009[a] | HAR0709C | CBF, Harichand | Cattle | HQ439215–JX171679[b] | Jamal et al. (2011a) This study |
| O/NWF/PAK/L1418/2009[a] | HAR0809C | CBF, Harichand | Cattle | HQ439216–JX171680[b] | Jamal et al. (2011a) This study |
| O/PAK/31/2005 | HBD0105C | Hafizabad | Cattle | FJ798162 | Waheed et al. (2011) |
| O/PAK/32/2005 | HBD0205C | Hafizabad | Cattle | FJ798163 | Waheed et al. (2011) |
| O/PAK/35/2005 | LHR0805B | Lahore | Buffalo | FJ798166 | Waheed et al. (2011) |
| O/PAK/33/2005 | AWA0905C | Arifwala | Cattle | FJ798164 | Waheed et al. (2011) |
| O/PAK/34/2005 | AWA1005C | Arifwala | Cattle | FJ798165 | Waheed et al. (2011) |
| O/PAK/60/2006 | SGD0606B | Sargodha | Buffalo | FJ798168 | Waheed et al. (2011) |

  

**Table 1.** Continued.

| Sequence name | Fig. 2 Nomenclature | Sample location | Species affected | Accession. number | Reference |
|---|---|---|---|---|---|
| O/PAK/63/2006 | SKP0706B | Sheikhopura | Buffalo | FJ798170 | Waheed et al. (2011) |
| O/PAK/72/2006 | GJW0806B | Gujranwala | Buffalo | FJ798176 | Waheed et al. (2011) |
| O/PAK/73/2006 | GJW0906B | Gujranwala | Buffalo | FJ798177 | Waheed et al. (2011) |
| O/PAK/74/2006 | OKA1006C | Okara | Cattle | FJ798178 | Waheed et al. (2011) |
| O/PAK/66/2006 | JHG2006C | Jhang | Cattle | FJ798171 | Waheed et al. (2011) |
| O/PAK/67/2006 | JHG2106B | Jhang | Buffalo | FJ798172 | Waheed et al. (2011) |
| O/PAK/70/2006 | AWA2206C | Arifwala | Cattle | FJ798174 | Waheed et al. (2011) |
| O/PAK/71/2006 | AWA2306C | Arifwala | Cattle | FJ798175 | Waheed et al. (2011) |
| O/PAK/60/2007 | JHG0207C | Jhang | Cattle | FJ798181 | Waheed et al. (2011) |
| O/PAK/56/2007 | SKP0307B | Sheikhopura | Buffalo | FJ798180 | Waheed et al. (2011) |
| O/PAK/61/2007 | HBD0407C | Hafizabad | Cattle | FJ798182 | Waheed et al. (2011) |
| O/PAK/63/2007 | GJW0507C | Gujranwala | Cattle | FJ798183 | Waheed et al. (2011) |
| O/PAK/66/2007 (A)[c] | LYH1007B | Layyah | Buffalo | FJ798184 | Waheed et al. (2011) |
| O/PUN/PAK/L1346/2008 | FBD0508C | Faisalabad | Cattle | HQ439209 | Jamal et al. (2011a) |
| O/PUN/PAK/L1347/2008 | FBD0608C | Faisalabad | Cattle | HQ439206 | Jamal et al. (2011a) |
| O/PUN/PAK/L1358/2008[a] | FBD0708C | Faisalabad | Cattle | HQ439208–JX171676[b] | Jamal et al. (2011a) |
| O/PUN/PAK/L1353/2008 | FBD0808C | Faisalabad | Cattle | HQ439211 | Jamal et al. (2011a) |
| O/PUN/PAK/L1360/2008 | FBD0908C | Faisalabad | Cattle | HQ439210 | Jamal et al. (2011a) |
| O/PUN/PAK/L1345/2008 | FBD1008B | Faisalabad | Buffalo | HQ439213 | Jamal et al. (2011a) |
| O/PUN/PAK/L1343/2008 | FBD1108C | Faisalabad | Cattle | HQ439207 | Jamal et al. (2011a) |
| O/ISL/PAK/L1412/2009[a] | IBD0109C | Islamabad | Cattle | HQ439214–JX171678[b] | Jamal et al. (2011a) |
| O/ISL/PAK/L1413/2009[a] | IBD0209C | Islamabad | Cattle | HQ439220–HQ113232[d] | Jamal et al. (2011a) Belsham et al. (2011) |
| O/KHI/PAK/04/2009[a] | KHI1109C | Landhi, Karachi | Cattle | JX170744 | This study |
| O/KHI/PAK/08/2009[a] | KHI1009B | Sargani, Karachi | Buffalo | JX170745 | This study |
| O/SGD/PAK/17/2010[a] | SGD0110B | Sargodha | Buffalo | JX170753 | This study |
| O/SGD/PAK/18/2010[a] | SGD0310C | Sargodha | Cattle | JX170754 | This study |
| O/SWL/PAK/16/2010[a] | SWL0210B | Sahiwal | Buffalo | JX170752 | This study |
| O/BWP/PAK/13/2010[a] | BWP0410B | Bahawalpur | Buffalo | JX170749 | This study |
| O/SGD/PAK/09/2010[a] | SGD0510C | Sargodha | Cattle | JX170746 | This study |
| O/IBD/PAK/10/2010[a] | IBD0610C | Islamabad | Cattle | JX170747 | This study |
| O/BWL/PAK/12/2010[a] | BWP0710B | Bahawalpur | Buffalo | JX170748 | This study |
| O/RWP/PAK/14/2010[a] | RWP0810B | Rawalpindi | Buffalo | JX170750 | This study |
| O/RWP/PAK/15/2010[a] | RWP0910B | Rawalpindi | Buffalo | JX170751 | This study |
| O/KHI/PAK/41/2011[a] | KRI0111B | Karachi | Buffalo | JX170756 | This study |
| O/KHI/PAK/42/2011[a] | KRI0211B | Karachi | Buffalo | JX170757 | This study |
| O/SGD/PAK/19/2011[a] | SGD0311C | Sargodha | Cattle | JX170755 | This study |

[a]All P1 sequence available.
[b]Accession number for P1 sequence.
[c]Not included in cluster analysis, sequences fell in cluster of n<3 sequences.
[d]P1 sequence derived from rescued virus.

the authors upon request. The P1 sequences were submitted to GenBank and assigned accession numbers JX170744–JX170757. Additional P1 sequences generated at DTU-VET, Lindholm, were generated as described previously (Jamal et al., 2011a), and the sequences have been assigned accession numbers JX171676–JX171680 (Table 1).

### VP1 and P1 nt and aa pair-wise difference correlation

To assess the hypothesis that VP1 alone (639 nt for the isolates listed in Table 1) contains information comparable to that provided by the entire P1 (2202 nt), we aligned the 20 P1 sequences using ClustalW algorithm (Larkin et al., 2007) implemented in BioEdit® (Hall, 1999). Subsequently, we computed all pair-wise nt and aa differences of P1 and their respective VP1. The nt and aa differences for VP1 between each pair relative to the whole P1 were computed as a percentage (#nt or aa differences in VP1/# nt or aa differences in P1). Additionally, Pearson's coefficient of correlation ($r$) for the nt and aa changes in VP1 and P1 was computed in PASW Statistics 18m release version 18.0.0 (SPSS, Inc., 2009, Chicago, IL, USA).

 

### Statistical model to assess the association of genetic variation and epidemiological factors

Genetic clusters of the 82 FMDV-O VP1 nucleotide sequences were identified by constructing a neighbour-joining tree using Phylip software package (Felsenstein, 2005) and visualized in Fig Tree v1.3.1 (Rambaut, 2006–2009). The F84 substitution model with gamma distributed rates across sites was used. Additionally, a phylogenetic tree using the predicted aa sequences was constructed to compare clustering of viruses based on aa and nt sequences.

A hierarchical linear model was formulated, where the nt and aa distance from each sequence to the earliest isolate in the cluster was the dependent variable. Host species, time difference and spatial distance between the earliest sample in the cluster and each of the remaining sequences in a cluster were assessed for fixed effects in the model. Time was calculated as the number of days between the sample collection, and spatial distance between the two sample point locations (latitude and longitude) was computed using the Haversine distance formula. Differences in species were categorized in samples collected from the same or different species (cattle and buffalo). The genetic cluster was used as a grouping variable (hierarchy). Models for prediction were selected based on the value of the squared correlation between predicted and observed values, to assess the fit of the mixed models, and by the Akaike information criterion (AIC). All statistical analyses were performed using PASW Statistics 18m release version 18.0.0 (SPSS, Inc.).

## Results

### VP1 and P1 nt and aa pair-wise difference correlation

The 20 whole P1 coding sequences were derived from samples collected from buffalo ($n = 11$) and cattle ($n = 9$) and originated in eight different locations. There were three pairs with identical nt sequences; interestingly, two of these identical samples were collected 1 day apart, from different species (buffalo and cattle) and in two different locations, Sargodha and Bahawalpur (Punjab Province), which are approximately 340 km apart from each other.

Pearson correlation coefficient of the distance measured in P1 and VP1 was high ($r = 0.99$, $P < 0.05$) for nucleotides and moderate ($r = 0.79$, $P < 0.05$) for the amino acid sequences (Fig. 1). The most divergent sequences had 50 and 137 nt differences for VP1 and P1, respectively, which resulted in five and 14 aa differences, respectively.

The mean proportion of VP1/P1 differences was 0.39 (SD = 0.12) and 0.68 (SD = 0.25) in nt and aa sequences, respectively.

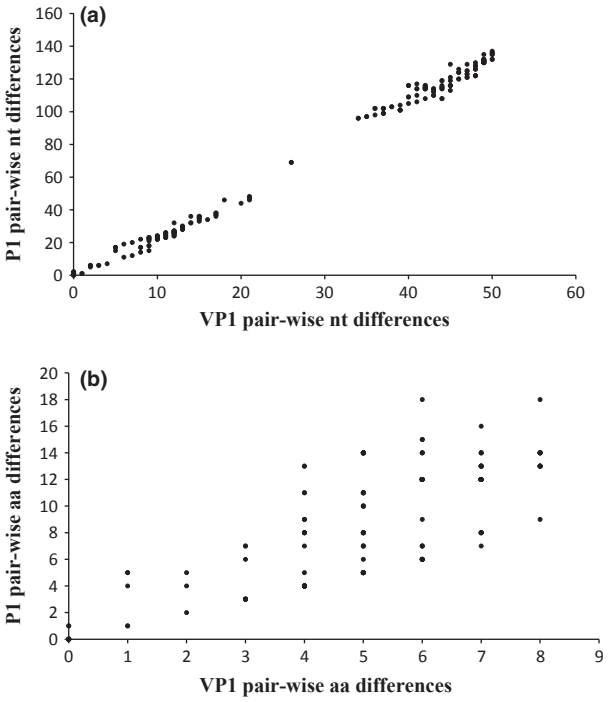


**Fig. 1.** Scatter plots of pair-wise differences between 20 VP1 and P1 nucleotide (a) and amino acid (b) sequences of FMDV serotype O collected from Pakistan.

### Statistical model to assess the association of genetic variation with epidemiological factors

The neighbour-joining tree of the 82 VP1 nucleotide sequences showed seven major genetic clusters with $n = 24$, 14, 8, 5, 7, 9 and 11 sequences, respectively (Fig. 2). Phylogenetic trees constructed using predicted aa sequences yielded similar clusters (not shown), so initial genetic clusters provided by nucleotide sequence analysis was considered an appropriate grouping criterion for models assessing nt and aa changes. Sequence variability within each cluster was <5% (≤30 nt difference). Sequence O/PAK/66/2007 (A) did not fall into any cluster and one of the clusters included only three sequences (O/PAK/38/2005, O/PUN/PAK/L282/2005 and O/ISL/PAK/L288/2005); therefore, those four sequences were not included in the regression model because of insufficient data for contributing to intracluster variability. Consequently, 79 VP1 sequences (36 from buffalo and 43 from cattle) collected at 22 different locations were included in the regression model (Fig. 2). These viruses all belong within the O PanAsia lineage, and all the additional viruses (collected in 2009–2011) sequenced in this study belonged to the PanAsia II$^{ANT-10}$ sublineage.

The model with the lowest AIC value included only spatial distance as a significant predictor for both the nt and aa

  

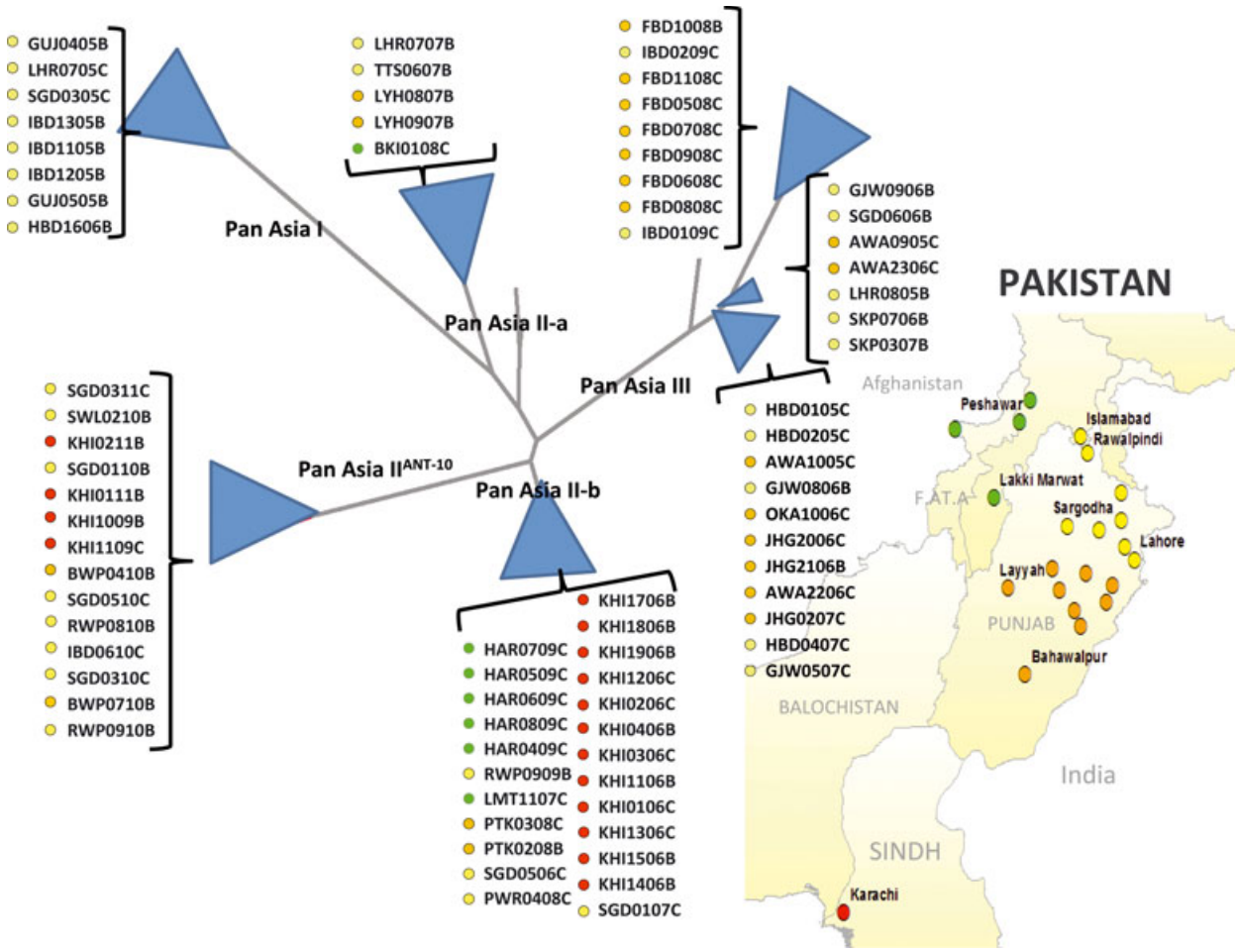


**Fig. 2.** Neighbour-joining tree of analysed VP1 nucleotide sequences. The topology of the tree shows seven major clusters with <5% difference within each cluster, while an eighth cluster was not included in additional analyses because of the low number of sequences within it ($n$ = 3). Previously named lineages within the O PanAsia viruses (see Jamal et al., 2011a) are also indicated in the tree branches. The sequences were named as follows: the first three text characters correspond to the lower administrative location of sample collection, the following two numbers are the consecutive samples from a particular year, two numbers indicate the year and the last capital letter denotes the species from which the sample was collected (C = cattle, B = buffalo). The ranges of the difference variables assessed in the model are given for each cluster. Sampling location of the sequences is given by the coloured circle in front of the sequence name, corresponding to circle location in the map.

models. The squared correlation between the predicted and the observed values was 0.83 and 0.87 for the nt and aa sequences, respectively. The model predicted that the VP1 nt sequences in the same location differed by 2.5 (95% CI = 1.0–4.1) nucleotides and differences increased by 1.7 (95%CI = 1.4–1.9) nucleotides per 100 km. Amino acid sequences in the same location differed by 0.9 (95%CI = 0.5–1.3) aa and differences increased by 0.5 (95% CI = 0.4–0.6) aa every 100 km. Host species was not significantly associated with aa or nt differences. Geographical distance and time between sampling were highly correlated ($r$ = 0.72, $P$ < 0.05), and for that reason, inclusion of time did not improve the model fitness.

## Discussion

Findings from this study provide information on the predictive nature of VP1, compared to P1, for assessing the nature and extent of nt and aa sequences variation of recent FMDV serotype O isolates from Pakistan. The 14 viruses sequenced for this analysis belonged to the PanAsia II$^{ANT-10}$ sublineage, which is currently dominant in the West Eurasia region and has spread as far west as Libya and Bulgaria. Two of these viruses were sampled in 2009, in Karachi dairy colonies, and are the earliest reported PanAsia II$^{ANT-10}$ viruses, suggesting that this sublineage may have originated in Pakistan.

Compared to P1 sequences, VP1 was a good predictor of nt differences, but its ability to predict aa changes was moderate. This finding may be related with a higher synonymous substitution accumulation throughout P1, but increased non-synonymous substitution in VP1, arguably, because of differential changes driven by selective pressure and/or allowed by structural constraints (Sobrino et al., 2001). These results suggest that at least for serotype O in Pakistan and possibly other serotypes in endemic settings, the entire P1 sequences may be required to assess aa variation and immunological studies such as vaccine matching, whereas VP1 sequence alone may be sufficient when the objective of the study is to assess nt variations.

Few earlier studies have attempted to associate epidemiological factors with genetic variation in FMDV. Pairwise VP1 genetic variation for the epidemic that affected Argentina in 2001 was associated with longer distances and greater time intervals and with the interaction between incidence and duration of outbreaks (Perez et al., 2008). In that study, the regression model did not account for genetic clustering because all isolates included were from a single epidemic. In a different study, no association between nt substitution rate and host species or country of origin was found in O PanAsia FMDVs collected from 2005 to 2009 in Pakistan and Afghanistan (Jamal et al., 2011a). This lack of geographical association could be due to lack of specific point location information, because samples were only classified at the country level. In another study, in which FMDV-O PanAsia lineage sequences collected from 1989 to 2001 in the Middle East, Asia, Europe and Africa were analysed, significant associations with host and region were found (Garabed et al., 2009). Contrasting findings were presented in a global study of FMDV serotypes O and A, sampled between 1939 and 2010, where no association with FMD evolution and time, country or host species was identified (Yoon et al., 2011a). The wide time range and geographical diversity of viruses included in the aforementioned epidemiological analysis may have resulted in low resolution (spatial and temporal) of the FMDV phylodynamics, preventing the detection of geographical or temporal associations.

The interpretation of all these epidemiological studies must take into account the fact that accumulated mutations observed in field samples over time are the result of the virus mutation rate and the selective forces exerted on the virus population in a given environment. Therefore, both the geographical diversity and temporal breadth of the samples studied can dramatically influence the epidemiological findings (Muellner et al., 2011).

In our study, we fitted a model that included genetic cluster as a hierarchy, to compute only the genetic distance to the earliest isolate in the cluster. In the final models, only geographical distance was significantly associated with aa and nt variation. However, when time between isolations replaced geographical distance, the association between time and aa or nt diversity was also significant (results not shown). Lack of time association with genetic variation was likely due to the strong correlation between geographical distance and time between isolation of the samples, which impairs our ability to understand whether sequences were different because they were collected at different locations or at different periods of time.

Interestingly, collection of samples from different species was not significantly associated with the pair-wise genetic distance assessed here. Moreover, we found three identical P1 nt sequences, two of them sampled 1 day apart, from buffalo and cattle, and in two different locations (Sargodha and Bahawalpur). These results support early findings, suggesting that FMDV-O Pan Asia lineage is adapted to multiple host species (Knowles et al., 2005) and is consistent with the frequent multi-species outbreaks observed in Pakistan and with the waves of disease outbreaks at specific locations and dates. These results have implications for disease prevention and control as it indicates transmission and circulation of similar viruses among different species.

The estimates given in this study add a geographical component to analysing genetic variability among field isolates, but they should be carefully interpreted as an average variation of the viruses studied. Some viruses may show lower or higher variation, having important implications to the disease epidemiology in endemic settings. For instance, in Fig. 2, most of the clusters are genetically diverse, except for the two smallest clusters of the Pan Asia III lineage, although they were collected within a time span of 761 and 694 days, respectively and within a spatial distance of 198 and 164 km. Only one nucleotide difference was observed in both clusters, resulting in one amino acid change in one of them. This sequence conservation for FMDV is unusual, given the fact that these viruses are interacting with hosts that have been either previously exposed or possibly vaccinated with FMDV. These conserved sequences of VP1 in space and time may suggest that these lineages use a mechanism for perpetuating in the animal population that does not require antigenic adaptation to host immune responses.

Arguably, the most important limitation of the study here was the impossibility to perform a probabilistic sampling design of outbreaks because of the nature and extent of the epidemiological information made available to us. Certainly, information on factors such as animal breed, age, gender, management practices, animals present in the herd, animal movement, details of the clinical disease and lesions would have helped to improve the model

  

fitness and to reveal unexplored associations and influences. Unfortunately, implementation of structured study designs and availability of information has been impaired here by the particular prevailing conditions of the settings, as is typically the case for observational studies in FMD endemic regions.

In conclusion, this study contributes to the knowledge of FMDV dynamics in space and time, which, in addition to a comprehensive understanding of the molecular and structural changes related to virus antigenicity, will help to design appropriate FMD control strategies in endemic settings.

## Acknowledgements

B. Brito and Z. Ahmed are recipients of a Plum Island Animal Disease Center Research Participation Program fellowship administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the U.S. Department of Agriculture (USDA). All opinions expressed in this paper are the author's and do not necessarily reflect the policies and views of the USDA, DOE, ORAU/ORISE. This research was funded in part by USDA ARS-Pakistan Foot-and-Mouth Disease project 58-1940-7-161F.

## References

Acharya, R., E. Fry, D. Stuart, G. Fox, D. Rowlands, and F. Brown, 1989: The three-dimensional structure of foot-and-mouth disease virus at 2.9 A resolution. *Nature* 337, 709–716.

Belsham, G. J., 2005: Translation and replication of FMDV RNA. *Curr. Top. Microbiol. Immunol.* 288, 43–70.

Belsham, G. J., S. M. Jamal, K. Tjørnehøj, and A. Bøtner, 2011: Rescue of foot-and-mouth disease viruses that are pathogenic for cattle from preserved viral RNA samples. *PLoS One* 6, e14621. doi:10.1371/journal.pone.0014621.

Brooksby, J. B., 1958: The virus of foot-and-mouth disease. *Adv. Virus Res.* 5, 1–37.

Carrillo, C., E. R. Tulman, G. Delhon, Z. Lu, A. Carreno, A. Vagnozzi, G. F. Kutish, and D. L. Rock, 2005: Comparative genomics of foot-and-mouth disease virus. *J. Virol.* 79, 6487–6504.

Domingo, E., C. Escarmis, E. Baranowski, C. M. Ruiz-Jarabo, E. Carrillo, J. I. Nunez, and F. Sobrino, 2003: Evolution of foot-and-mouth disease virus. *Virus Res.* 91, 47–63.

Felsenstein, J., 2005: PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5, 164–166.

Garabed, R. B., W. O. Johnson, and M. C. Thurmond, 2009: Analytical epidemiology of genomic variation among Pan Asia strains of foot-and-mouth disease virus. *Transbound. Emerg. Dis.* 56, 142–156.

Grubman, M. J., and B. Baxt, 2004: Foot-and-mouth disease. *Clin. Microbiol. Rev.* 17, 465–493.

Hall, T. A., 1999, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser. 41*, 95–98.

Haydon, D. T., A. D. Bastos, N. J. Knowles, and A. R. Samuel, 2001: Evidence for positive selection in foot-and-mouth disease virus capsid genes from field isolates. *Genetics* 157, 7–15.

Jamal, S. M., S. Ahmed, M. Hussain, and Q. Ali, 2010: Status of foot-and-mouth disease in Pakistan. *Arch. Virol.* 155, 1487–1491.

Jamal, S. M., G. Ferrari, S. Ahmed, P. Normann, and G. J. Belsham, 2011a: Genetic diversity of foot-and-mouth disease virus serotype O in Pakistan and Afghanistan, 1997-2009. *Infect. Genet. Evol.* 11, 1229–1238.

Jamal, S. M., G. Ferrari, S. Ahmed, P. Normann, and G. J. Belsham, 2011b: Molecular characterization of serotype Asia-1 foot-and-mouth disease viruses in Pakistan and Afghanistan; emergence of a new genetic Group and evidence for a novel recombinant virus. *Infect. Genet. Evol.* 11, 2049–2062.

Jamal, S. M., G. Ferrari, S. Ahmed, P. Normann, S. Curry, and G. J. Belsham, 2011c: Evolutionary analysis of serotype A foot-and-mouth disease viruses circulating in Pakistan and Afghanistan during 2002-2009. *J. Gen. Virol.* 92, 2849–2864.

James, A. D., and J. Rushton, 2002: The economics of foot and mouth disease. *Rev. Sci. Tech.* 21, 637–644.

Klein, J., M. Hussain, M. Ahmad, M. Afzal, and S. Alexandersen, 2008: Epidemiology of foot-and-mouth disease in Landhi Dairy Colony, Pakistan, the world largest Buffalo colony. *Virol. J.* 5, 53.

Knowles, N. J., and A. R. Samuel, 2003: Molecular epidemiology of foot-and-mouth disease virus. *Virus Res.* 91, 65–80.

Knowles, N. J., A. R. Samuel, P. R. Davies, R. J. Midgley, and J. F. Valarcher, 2005: Pandemic strain of foot-and-mouth disease virus serotype O. *Emerg. Infect. Dis.* 11, 1887–1893.

Larkin, M. A., N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. Higgins, 2007: ClustalW and ClustalX version 2. *Bioinformatics* 23, 2947–2948.

Muellner, P., R. Zadoks, A. Perez, S. E. F. Spencer, Y. H. Schukken, and N. P. French, 2011: The integration of molecular tools into veterinary and spatial epidemiology. *Spat. Spatiotemporal Epidemiol.* 2, 159–171.

Pacheco, J. M., J. Arzt, and L. L. Rodriguez, 2010: Early events in the pathogenesis of foot-and-mouth disease in cattle after controlled aerosol exposure. *Vet. J.* 183, 46–53.

Pauszek, S. J., C. Barrera Jdel, T. Goldberg, R. Allende, and L. L. Rodriguez, 2011: Genetic and antigenic relationships of vesicular stomatitis viruses from South America. *Arch. Virol.* 156, 1961–1968.

Perez, A. M., G. Konig, E. Spath, and M. C. Thurmond, 2008: Variation in the VP1 gene of foot-and-mouth disease virus serotype A associated with epidemiological characteristics of outbreaks in the 2001 epizootic in Argentina. *J. Vet. Diagn. Invest.* 20, 433–439.

 

Perry, B. D., and K. M. Rich, 2007: Poverty impacts of foot-and-mouth disease and the poverty reduction implications of its control. *Vet. Rec.* 160, 238–241.

Rambaut, A.. 2006–2009. *Fig Tree. Tree Figure Drawing Tool, Version 1.3.1*. Available at http://tree.bio.ed.ac.uk/software/figtree/ (accessed July 16, 2012).

Sobrino, F., M. Saiz, M. A. Jimenez-Clavero, J. I. Nunez, M. F. Rosas, E. Baranowski, and V. Ley, 2001: Foot-and-mouth disease virus: a long known virus, but a current threat. *Vet. Res.* 32, 1–30.

Thompson, D., P. Muriel, D. Russell, P. Osborne, A. Bromley, M. Rowland, S. Creigh-Tyte, and C. Brown, 2002: Economic costs of the foot and mouth disease outbreak in the United Kingdom in 2001. *Rev. Sci. Tech.* 21, 675–687.

Valarcher, J. F., Y. Leforban, M. Rweyemamu, P. L. Roeder, G. Gerbier, D. K. Mackay, K. J. Sumption, D. J. Paton, and N. J. Knowles, 2008: Incursions of foot-and-mouth disease virus into Europe between 1985 and 2006. *Transbound. Emerg. Dis.* 55, 14–34.

Valarcher, J. F., N. J. Knowles, V. Zakharov, A. Scherbakov, Z. Zhang, Y. J. Shang, Z. X. Liu, X. T. Liu, A. Sanyal, D. Hemadri, C. Tosh, T. J. Rasool, B. Pattnaik, K. R. Schumann, T. R. Beckham, W. Linchongsubongkoch, N. P. Ferris, P. L. Roeder, and D. J. Paton, 2009: Multiple origins of foot-and-mouth disease virus serotype Asia 1 outbreaks, 2003-2007. *Emerg. Infect. Dis.* 15, 1046–1051.

Waheed, U., S. Parida, Q. M. Khan, H. Hussain, K. Ebert, J. Wadsworth, G. H. S. M. Hutchings, M. Mahapatra, D. P. King, D. J. Paton, and N. J. Knowles, 2011: Molecular characterisation of foot-and-mouth disease viruses from Pakistan, 2005–2008. *Transbound. Emerg. Dis.* 58, 166–172.

Yoon, S. H., K. N. Lee, J. H. Park, and H. Kim, 2011a: Molecular epidemiology of foot-and-mouth disease virus serotypes A and O with emphasis on Korean isolates: temporal and spatial dynamics. *Arch. Virol.* 156, 817–826.

Yoon, S. H., W. Park, D. P. King, and H. Kim, 2011b: Phylogenomics and molecular evolution of foot-and-mouth disease virus. *Mol. Cells* 31, 413–421.